

## Coronavirus Pandemic

# Non-linear link between temperature difference and COVID-19: Excluding the effect of population density

Yongmei Ding<sup>1</sup> #, Liyuan Gao<sup>1</sup> #, Ning-Yi Shao<sup>2</sup>

<sup>1</sup> Department of Mathematics and Statistics, College of Science, Wuhan University of Science and Technology, Wuhan, Hubei Province, China

<sup>2</sup> Faculty of Health Sciences, University of Macau, Macau SAR, China

# Authors contributed equally, and should be regarded as co-first authors.

### Abstract

**Introduction:** The spatiotemporal patterns of Corona Virus Disease 2019 (COVID-19) is detected in the United States, which shows temperature difference (TD) with cumulative hysteresis effect significantly changes the daily new confirmed cases after eliminating the interference of population density.

**Methodology:** The nonlinear feature of updated cases is captured through Generalized Additive Mixed Model (GAMM) with threshold points; Exposure-response curve suggests that daily confirmed cases is changed at the different stages of TD according to the threshold points of piecewise function, which traces out the rule of updated cases under different meteorological condition.

**Results:** Our results show that the confirmed cases decreased by 0.390% (95% CI: -0.478~ -0.302) for increasing each one degree of TD if TD is less than 11.5°C; It will increase by 0.302% (95% CI: 0.215 ~ 0.388) for every 1°C increase in the TD (lag0-4) at the interval [11.5, 16]; Meanwhile the number of newly confirmed COVID-19 cases will increase by 0.321% (95% CI: 0.142 ~ 0.499) for every 1°C increase in the TD (lag0-4) when the TD (lag0-4) is over 16°C, and the most fluctuation occurred on Sunday. The results of the sensitivity analysis confirmed our model robust.

**Conclusions:** In US, this interval effect of TD reminds us that it is urgent to control the spread and infection of COVID-19 when TD becomes greater in autumn and the ongoing winter.

**Key words:** Coronavirus; temperature difference; nonlinearity; population density.

*J Infect Dev Ctries* 2021; 15(2):230-236. doi:10.3855/jidc.13926

(Received 16 September 2020 – Accepted 25 January 2021)

Copyright © 2021 Ding *et al.* This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Introduction

The outbreak of Corona Virus Disease 2019 (COVID-19), which poses a public health threat [1,2], has led to more than 93 million subjected infected and 2.01 million dead worldwide as of 17 January 2021 [3], correspondingly up to 23.34 million and 0.39 million for the United States (US), respectively. Early studies have shown that COVID-19 is transmitted between individuals by direct contact or droplets, like coughing, sneezing, talking, or even singing [4]. Air pollution and population density have also been proved as important factors for the transmission and survival of this coronavirus [5-9]. Meteorological changes in indoor and outdoor environmental factors on human behavior, social interactions, or hygiene practices that stimulate the propagation among the infected or susceptible [10]. However, less is certain about the environmental conditions that drive the spatiotemporal patterns of COVID-19 in the US as the confirmed cases are increased dramatically day by day, especially since 26

June 2020, the number of daily new cases is over 40 thousand; even more, up to 200 thousand in 20 December 2020. Some scientists conjecture that low temperature regions enhance the viability of COVID-19 cases [11-13]. However, other researchers deem daily average temperature is correlated positively with the number of daily increased COVID-19 cases or weak association between them [14,15]. Here we modeled the effect of temperature difference on COVID-19 transmission using a moving average lag period of 7 days. It also provides the inflection point of all orders of a lag period after excluding the impact of population density on newly confirmed cases. Nonlinear characteristics of daily confirmed cases are absorbed by GAMM, which provides the exact measure to spatiotemporal meteorological data. The influence of meteorological parameters is measured independently by the models' coefficients, and it is significant on the Daily New Cases (DNC). The sensitivity analysis is

implemented for the effect, and the consistency shows our models robust.

### Methodology

#### Data feature

Our case study covered 9 main partitions, including all 50 states and Washington, D.C in the US, the time span is from March 9th, 2020 to January 10, 2021. The data set information is extracted from the center of Systems Science and Engineering, Johns Hopkins University, while meteorological data is from the National Aeronautics and Space Administration. Figure 1 and Figure 2 show the cumulative confirmed cases and temperature differences in different states as of January 10, 2021. The darker the heat map, the bigger the data. Both cumulative confirmed cases of COVID-19 (Figure 1) and temperature changes (Figure 2) in the western and southern United States were significantly higher than in other regions.

The dependent variable is COVID-19 DNC. It is known that population density is an influenced factor to the increased cases of COVID-19 [16-18], with correlation coefficient of 0.265. The density coefficient

Figure 1. Cumulatively confirmed cases on January 10, 2021.

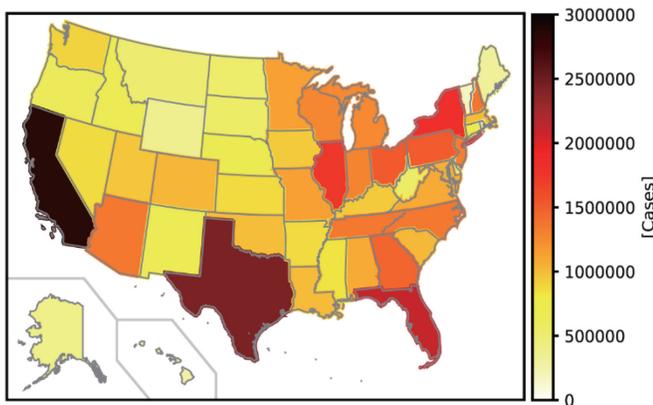


Figure 2. Temperature difference on January 10, 2021.

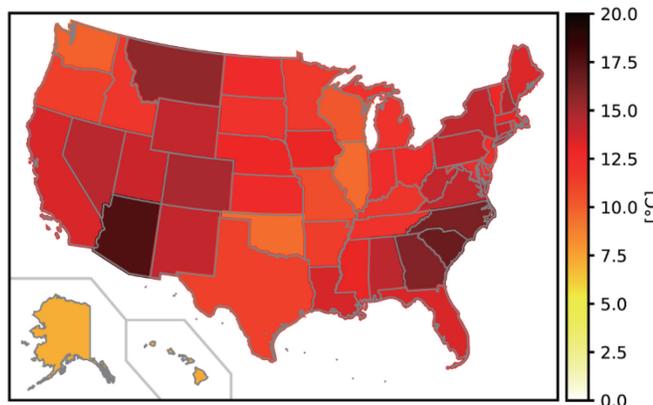


Table 1. Variables and their connotation.

Variables	Connotation
DNC	Daily new cases
TD	Temperature difference
ADR	Average daily rainfall rate
RH	The daily average of relative humidity
PS	The daily average of atmospheric pressure at the surface of the earth
WS	The daily average range of wind speed

of standardization for population density is derived by z-score. Here we eliminate the effect of population density through the COVID-19 confirmed cases dividing by this coefficient. The meteorological factors that might be associated with COVID-19 in the US, is shown in Table 1. The abbreviation forms are given as shown in the first column.

#### Variance Inflation Factors (VIF)

When multiple variables are analyzed, information overlap often exists between variables. Collinearity test is a suitable way to check the repeated information by the statistic of VIF with the expression shown in (1).  $R^2$  is the complex coefficient of regression to other independent variables with  $X(\beta)$  as the dependent variable. Usually, the multiple correlations will seriously affect the least-squares estimation if the maximum VIF exceeds 10 [19,20].

$$VIF(\hat{\beta}) = \frac{1}{(1-R^2)} \tag{1}$$

#### Generalized additive mixed model (GAMM)

Linear regression is often used when looking for correlations between multiple variables. This is not always the case. Based on the VIF test, we found that the nonlinear feature is more significant about the association with meteorological factors and COVID-19 cases. GAMM is used to explore the nonlinear relationship between weather factors and health conditions [21,22]. Since temperature may last for several days and the incubation period of COVID-19 varies from 1 to 14 days, we used GAMM to test the moving average hysteresis effect of temperature differences (lag0-4). The model is defined as follows:

$$\text{Log}(DNC_{it}) = \alpha + \beta_0(WEEK_i) + \beta_1(STATE_i) + s(TD_{it}, k = 5) + s(ADR_{it}, k = 3) + s(PS_{it}, k = 4) + s(WS_{it}, k = 3) + \varepsilon_{it} \tag{2}$$

where  $DNC_{it}$  denotes the number of new confirmed cases in the area  $i$  on day  $t$  (plus 1 to avoid the logarithm of 0) [23];  $\alpha$  is the intercept of model,  $WEEK_i$  and  $STATE_i$  are named as indicative functions to denote the effect of week and state for controlling the effects of

short-term fluctuations in time and random effects in different regions.  $s(\cdot)$  is a natural spline function for smoothing, and the degree of freedom is selected according to Akaike information criterion (AIC) to avoid overfitting [24].

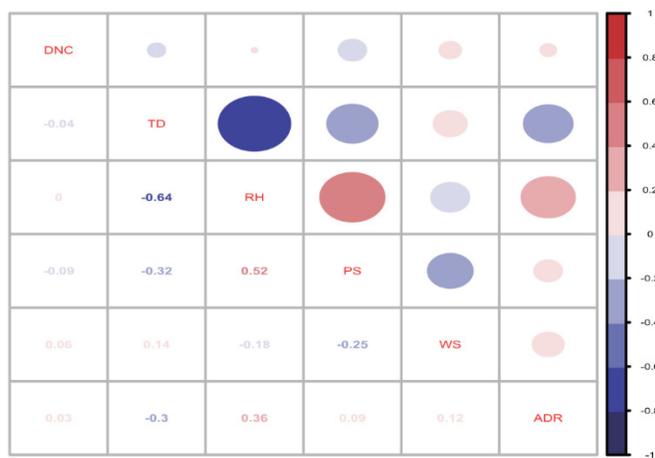
**Results**

*Descriptive analysis*

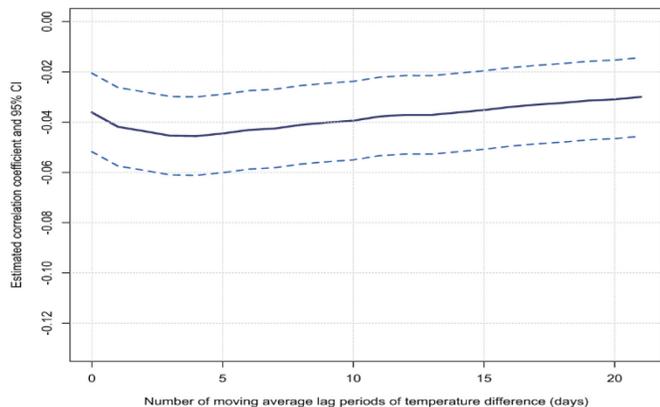
It shows the correlation coefficients between meteorological variables in Figure 3. The heat map shows there exists a strong correlation between RH and the four factors of TD & ADR & PS & WS, which means the latter variables include the most information of RH, so we haven't selected it in our modeling process to avoid overfitting in the following section.

The TD is a significant factor to affect the transmission of COVID-19 cases, but we cannot sure

**Figure 3.** Thermal diagram of correlation coefficient between meteorological variables.



**Figure 4.** Diagram of correlation coefficient between DNC and TD (lag0 ~ lag0-21).



Solid line represents estimated correlation coefficient; dashed line represents 95% confidence interval of estimated correlation coefficient.

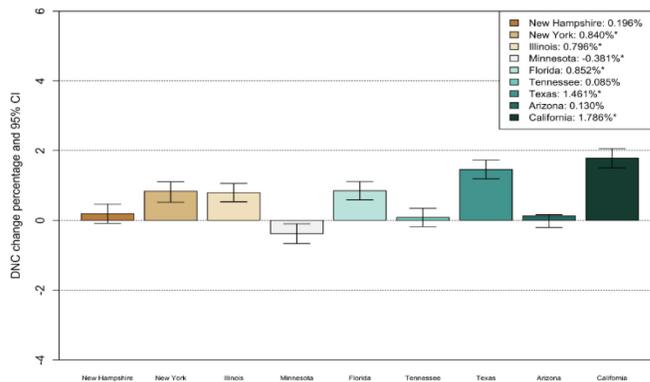
when is the most significant to this association. Different leading time is considered because of the incubation period of COVID-19, and the average effect of TD lag from 0 to 21 days is considered. The changing curve of this association is show in Figure 4. It seems smoothly, but the inflection point is emerged at the lag order of 4th.

*The association with daily updates of COVID-19 and TD*

We test the multicollinearity about the variables from Figure 3, the value of VIF shows no co-linear effect between them, which means that little evidence could express linear correlation among factors. Using the GAMM model, we found that when the moving average of TD lags for 4 days, the nonlinear relationship is robust and significant ( $p < 0.05$ ).

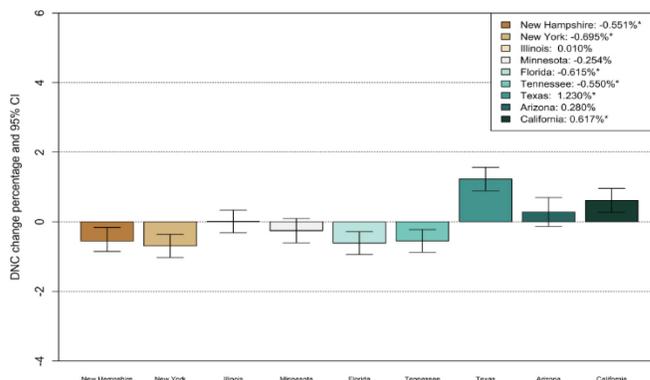
The space difference feature has aroused our attention. Nine representative states are selected according to the Census Bureau's Geographical

**Figure 5.** The percentage change of newly confirmed COVID-19 cases (lag0-4) for every 1°C change in TD before considering population density.



\*:  $p < 0.05$ , the same below.

**Figure 6.** For every 1°C change in TD after considering population density, the percentage change of newly confirmed COVID-19 cases (lag0-4).



Partition for detecting this difference. Here we demonstrate the difference of TD effect in Figure 5 and Figure 6. It also verifies the new confirmed cases have been posed significant influence by population density.

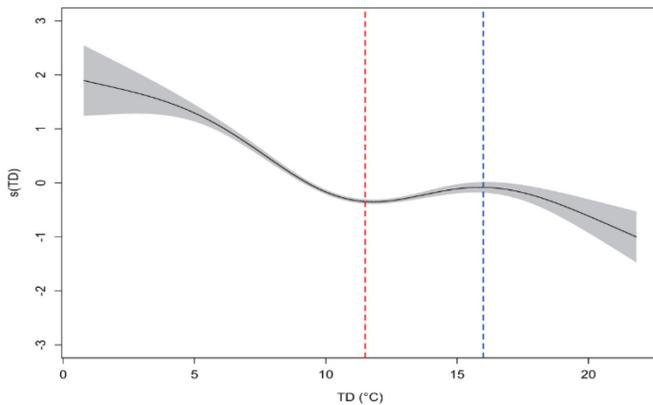
The association of nonlinear feature is evident from the exposure-response curve in Figure 7, which shows that the trend of newly confirmed COVID-19 cases influenced by TD (lag 0-4) is divided into three stages ( $p < 0.05$ ). Specifically, the relationship is approximately inversely linear when the temperature difference is less than 11.5°C or greater than 16°C, and positively linear when greater than 11.5°C and less than 16°C, which hints that the double thresholds of

temperature difference for new confirmed COVID-19 cases are set as 11.5°C and 16°C tentatively.

The association of COVID-19 cases with the combined effect of temperature difference and week effects is demonstrated in Figure 8. We try to find out the most significant week-time node, and the surface shows newly confirmed cases are waved at different range of TD and week time. It fluctuates the most significance on Sunday with the maximum change of -0.264% (95%CI: -0.381~ -0.381). The two thresholds of 11.5°C and 16°C make the effect of temperature difference into three parts. We chose linear model to quantify the impact of TD for every part, then a piecewise linear regression model is established here.

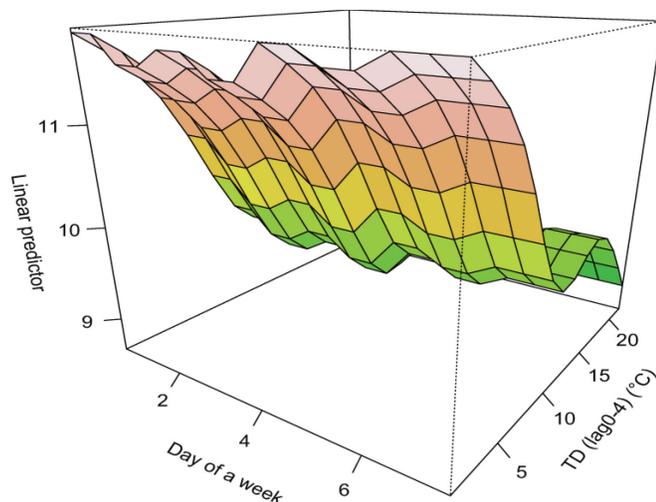
Our results show that the number of new confirmed COVID-19 cases decreases by 0.390% (95% CI: -0.478~ -0.302) for every 1°C increase when the TD (lag0-4) is lower than 11.5°C; it will increase by 0.302% (95% CI: 0.215 ~ 0.388) for every 1°C increase of TD (lag0-4) at the interval [11.5, 16], meanwhile the number of new confirmed cases will increase by 0.321% (95% CI: 0.142 ~ 0.499) for every 1°C increase when the TD (lag0-4) is over 16°C (Figure 9).

**Figure 7.** The exposure-response curve of the influence of temperature difference (lag0-4) on newly confirmed COVID-19 cases.

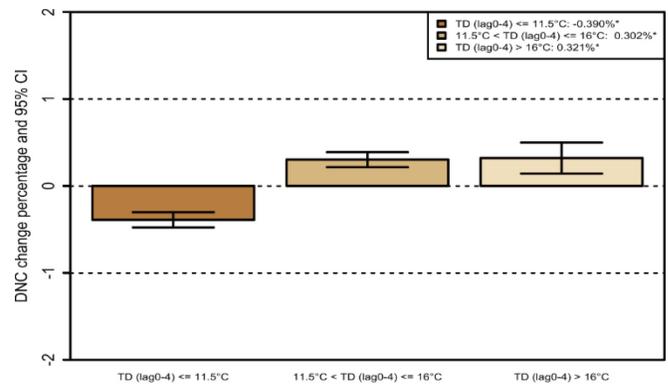


The X-axis is temperature difference, the Y-axis is the contribution of smoothness to the fitting value, and the red and blue dotted lines are the double thresholds of TD: 11.5°C and 16°C, respectively.

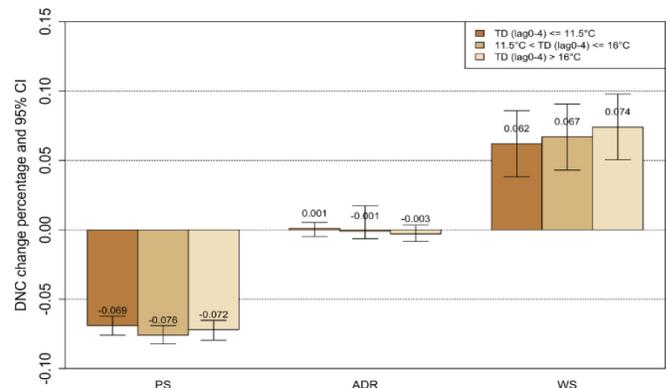
**Figure 8.** Perspective of new confirmed cases of COVID-19 affected by two variables of TD (lag0-4) and WEEK.



**Figure 9.** The role of TD (lag0-4) in the results of piecewise linear regression.



**Figure 10.** The role of other important meteorological variables (PS, ADR, WS) in the results of piecewise linear regression.

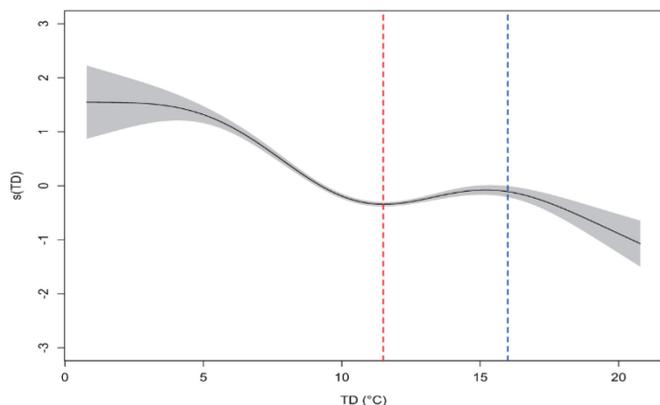


In the modeling process of piecewise linear regression, PS, ADR and WS are considered as the mixed effect to the updates on COVID-19 (Figure 10), which is highly correlated with TD. The weak effect is demonstrated through the coefficients in Figure 10. In the three stages of TD variation, the impact from these three variables is less than TD as the coefficients are almost lower than 0.1 with significances.

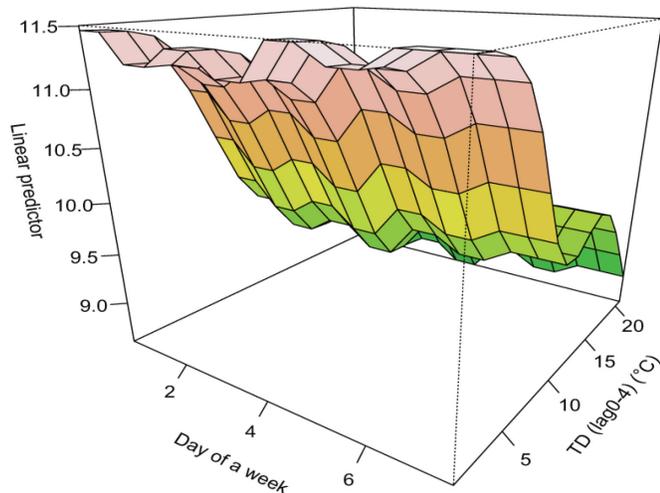
*Sensitivity analysis*

It is well known that California State is the hardest hit at the early time in the US, with a much larger cumulative number of confirmed cases than other cities. Here we exclude it from the data set for the sensitivity test. Our results show that the nonlinear relationship is not only robust but also significant when the moving average of TD is taken to lag by 4 days ( $p < 0.05$ ) (Figure 11).

**Figure 11.** The exposure-response curve of the influence of temperature difference (lag0-4) on newly confirmed COVID-19 cases.



**Figure 12.** Perspective of new confirmed cases of COVID-19 affected by two variables of TD (lag0-4) and WEEK.

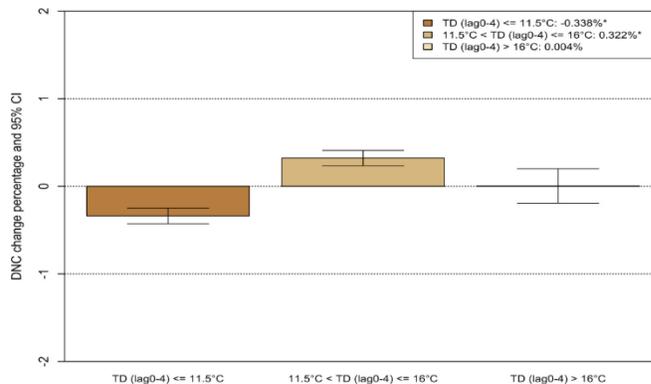


The number of new confirmed COVID-19 cases decreases by 0.338% (95%CI: -0.428~ -0.249) for every 1°C decreases if TD (lag0-4) is lower than the threshold of 11.5°C, while increases by 0.322% (95%CI: 0.235~ 0.410) if the TD (lag0-4) is greater than 11.5°C and less than 16°C, and increases by 0.004% (95% CI: -0.192 ~ 0.201) when the TD is higher than 16°C (Figure 13). The association of COVID-19 cases with the combined effect of temperature difference and week effects is demonstrated in Figure 12. In addition, Figure 14 verifies the conclusion that PS, ADR and WS have a weak impact on the COVID-19 update.

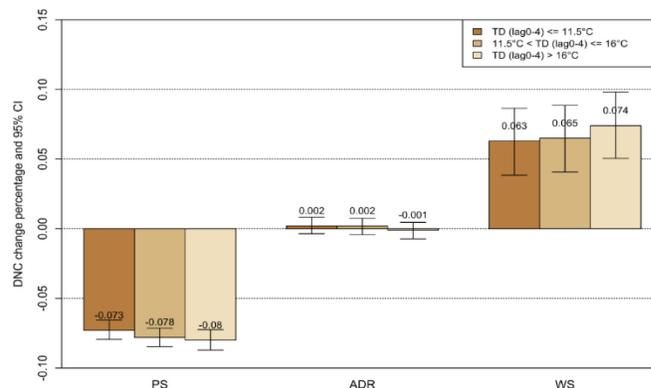
**Discussion**

We focus on the association of environmental temperature differences and newly confirmed cases of COVID-19. The nonlinear relationship is measured by GAMM, which is significant in the sensitive test. The greater the TD, the more updated cases of COVID-19 if the temperature difference is greater than 11.5°C and less than 16°C. When the temperature difference is greater than 16°C, the COVID-19 updates will further

**Figure 13.** The role of TD (lag0-4) in the results of piecewise linear regression.



**Figure 14.** The role of other important meteorological variables (PS, ADR, WS) in the results of piecewise linear regression.



increase, indicating that a higher TD may not inhibit the spread of this novel Coronavirus. But if the TD is lower than 11.5°C, it presents a reverse linear relationship that the greater the temperature difference, the fewer updates of COVID-19. External temperature changes that affect the transmission mechanism of COVID-19 mainly reflect in two aspects: one is the impact of temperature on the survival of COVID-19 virus [25-27], which has been confirmed in early studies on SARS virus [28-31]; second, external temperature changes cause the migration and mobility of population [32-34]. Our study sheds some light on the non-linear relationship between ambient temperature and new confirmed cases of COVID-19, which verifies that the number of new confirmed cases of COVID-19 may increase daily without any public health interventions if the weather changes dramatically. Therefore, neither the public nor the government can ignore the link between temperature changes and the virus and take preventive measures.

However, more impact factors may be detached to impact the new confirmed cases, which will be our further research points. First, public health intervention is an important factor in COVID-19 transmission. Second, our study only covered cities in the United States, further verification would be needed to other cities and regions.

### Acknowledgements

The authors wish to thank the study participants for their contribution to the research, as well as current and past investigators and staff.

### Authors' contributions

YD: Conceptualization, Methodology, Writing-Reviewing and Editing. LG: Software, Data curation, Writing-Original draft preparation, Visualization. NS: Supervision, Investigation.

### References

1. Saba A, Elsheikh A (2020) Forecasting the prevalence of COVID-19 outbreak in Egypt using nonlinear autoregressive artificial neural networks. *Process Safety and Environmental Protection* 141: 1-8.
2. Huang R, Liu M, Ding Y (2020) Spatial-temporal distribution of COVID-19 in China and its prediction: A data-driven modeling analysis. *J Infect Dev Ctries* 14: 246-253. doi: 10.3855/jidc.12585.
3. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z, Yu T, Xia J, Wei Y, Wu W, Xie X, Yin W, Li H, Liu M, Xiao Y, Gao H, Guo L, Xie J, Wang G, Jiang R, Gao Z, Jin Q, Wang J, Cao B (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 395: 497-506.
4. Lotfifi M, Hamblin M, Rezaei N (2020) COVID-19: Transmission, prevention, and potential therapeutic opportunities. *Clin Chim Acta* 508: 254-266.
5. Domingo J, Marqus M, Rovira J (2020) Influence of airborne transmission of sars-cov-2 on COVID-19 pandemic. A review. *Environ Res* 188: 109861.
6. Zoran M, Savastru R, Savastru DM, Tautan M (2020) Assessing the relationship between surface levels of pm2.5 and pm10 particulate matter impact on COVID-19 in Milan, Italy. *Sci Total Environ* 738: 139825.
7. Oztig LI, Askin OE (2020) Human mobility and coronavirus disease 2019 (COVID-19): a negative binomial regression analysis. *Public Health* 185: 364-367.
8. Iqbal N, Fareed Z, Shahzad F, He X, Shahzad U, Lina M (2020) The nexus between COVID-19, temperature and exchange rate in Wuhan city: New findings from partial and multiple wavelet coherence. *Sci Total Environ* 729: 138916.
9. Auler AC, Cássaro FAM, da Silva VO, Pires LF (2020) Evidence that high temperatures and intermediate relative humidity might favor the spread of COVID-19 in tropical climate: A case study for the most affected Brazilian cities. *Sci Total Environ* 729: 139090.
10. Runkle JD, Sugg MM, Leeper RD, Rao Y, Jessica L, Matthews, Jared JR (2020) Short-term effects of specific humidity and temperature on COVID-19 morbidity in select US cities. *Sci Total Environ* 740: 140093.
11. Wu Y, Jing W, Liu J, Ma Q, Yuan J, Wang Y, Du M, Liu M (2020) Effects of temperature and humidity on the daily new cases and new deaths of COVID-19 in 166 countries. *Sci Total Environ* 729: 139051.
12. Zhang Z, Xue T, Jin X (2020) Effects of meteorological conditions and air pollution on COVID-19 transmission: Evidence from 219 Chinese cities. *Sci Total Environ* 741: 140244.
13. Ujiie M, Tsuzuki S, Ohmagari N (2020) Effect of temperature on the infectivity of COVID-19. *Int J Infect Dis* 95: 301-303.
14. Menebo M (2020) Temperature and precipitation associate with COVID-19 new daily cases: A correlation study between weather and COVID-19 pandemic in Oslo, Norway. *Sci Total Environ* 737: 139659.
15. Bashir M, Ma B, Bilal, Komal B, Bashir M, Tan D, Bashir M (2020) Correlation between climate indicators and COVID-19 pandemic in New York, USA. *Sci Total Environ* 728: 138835.
16. Hamidi S, Ewing R, Sabouri S (2020) Longitudinal analyses of the relationship between development density and the COVID-19 morbidity and mortality rates: Early evidence from 1,165 metropolitan counties in the United States. *Health Place* 64: 102378.

17. Copiello S, Grillenzoni C (2020) The spread of 2019-ncov in china was primarily driven by population density. Comment on “Association between short-term exposure to air pollution and COVID-19 infection: Evidence from China by Zhu *et al.* *Sci Total Environ* 744: 141028.
18. Jahangiri M, Jahangiri M, Najafgholipour M (2020) The sensitivity and specificity analyses of ambient temperature and population size on the transmission rate of the novel coronavirus (COVID-19) in different provinces of Iran. *Sci Total Environ* 728: 138872.
19. Marquardt DW (1970) Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics* 12: 591-612.
20. Chennamaneni P, Echambadi R, Hess J, Syam N (2015) Diagnosing harmful collinearity in moderated regressions: A roadmap. *International Journal of Research in Marketing* 33: 172-182.
21. Cabrera M, Taylor G (2018) Modelling spatio-temporal data of dengue fever using generalized additive mixed models. *Spat Spatiotemporal Epidemiol* 28: 1-13.
22. Gaston RT, Ramroop S (2020) Prevalence of and factors associated with malaria in children under five years of age in Malawi, using Malaria indicator survey data. *Heliyon* 6: e03946.
23. Xie J, Zhu Y (2020) Association between ambient temperature and COVID-19 infection in 122 cities from china. *Sci Total Environ* 724: 138201.
24. Li W (2016) Analysis of the impact of major air pollutants on the deaths of residents in beijing and their spatial differences. Doctorate thesis in Epidemiology and Health Statistics. Peking Union Medical College. 23 p.
25. Mandal C, Panwar M (2020) Can the summer temperatures reduce COVID-19 cases? *Public Health* 185: 72-79.
26. Briz-Redón A, Serrano-Aroca A (2020). A spatio-temporal analysis for exploring the effect of temperature on COVID-19 early evolution in Spain. *Sci Total Environ* 728: 138811.
27. Alnefaie A, Albogami S (2020) Current approaches used in treating COVID-19 from a molecular mechanisms and immune response perspective. *Saudi Pharm J* 2: 1333- 1352.
28. Chan K, Peiris J, Lam S, Poon L, Seto W (2011) The effects of temperature and relative humidity on the viability of the SARS Coronavirus. *Adv Virol* 7: 734690.
29. Guionie O, Courtillon C, Allee C, Maurel S, Queguiner M, Eterradosi N (2013) An experimental study of the survival of turkey coronavirus at room temperature and + 4°C. *Avian Pathol* 42: 248-252.
30. Lowen A, Palese P (2009) Transmission of influenza virus in temperate zones is predominantly by aerosol, in the tropics by contact: a hypothesis. *Plos Curr* 1: RRN1002.
31. Ding Y, Gao L (2020) An evaluation of COVID-19 in Italy: A data-driven modeling analysis. *Infect Dis Model* 5: 495-501.
32. Tan J, Mu L, Huang J, Yu S, Chen B, Yin J (2005) An initial investigation of the association between the sars outbreak and weather: with the view of the environmental temperature and its variation. *J Epidemiol Community Health* 59: 186–192.
33. Shi P, Dong Y, Yan H, Zhao C, Li X, Liu W, He M, Tang S, Xi S (2020) Impact of temperature on the dynamics of the COVID-19 outbreak in China. *Sci Total Environ* 728: 138890.
34. Méndez-Arriaga F (2020) The temperature and regional climate effects on communitarian COVID-19 contagion in Mexico throughout phase I. *Sci Total Environ* 735: 139560.

### Corresponding authors

Yongmei Ding, PhD

Faculty of Statistics - Department of Mathematics and Statistics –  
Wuhan University of Science and Technology  
No.22317, 2nd Building, Huangjiahu Campus,  
Huangjiahu West Road, Hongshan District, Wuhan City, Hubei  
Province, P.R. of China, 430065  
Tel: 1-806-258-9916  
Fax: +86-027-68893261  
Email: dingyongmei@wust.edu.cn

Ning-Yi Shao, MBBS, PhD

Faculty of Health Sciences, University of Macau  
3009 Room, E12 Building  
Macau SAR, P.R. of China  
Tel: +853 8822-4895  
Fax: +853 8822 2314  
Email: nshao@um.edu.mo

**Conflict of interests:** No conflict of interests is declared.